

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:27:13 ; Search time 23.125 Seconds
(without alignments)
7069.698 Million cell updates/sec

Title: US-09-126-816B-6_COPY_1_1700

Perfect score: 8677

Sequence: 1 MNLVNAQLQKQWVYKFRIQ.....YLYGIDRVYKVIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76.*

1: p1r1.*

2: p1r2.*

3: p1r3.*

4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8646	99.6	2364	2 I40884	cytoxin L - Clostridium sordellii
2	6733.5	77.6	2366	2 S10317	toxin B - Clostridium sordellii
3	6711	77.3	2367	2 S70172	toxin A - Clostridium sordellii
4	4458	51.4	2710	2 A37052	alpha-toxin - Clostridium sordellii
5	2108.5	24.3	2178	2 S58005	toxin B - Escherichia coli
6	496.5	5.7	3169	2 T00296	rhoapry protein - Clostridium sordellii
7	470.5	5.4	2401	2 T28676	rhoapry protein - Clostridium sordellii
8	437	5.0	2269	2 T28677	rhoapry protein - Clostridium sordellii
9	424	4.9	2166	2 G70163	hypothetical protein
10	423.5	4.9	4688	2 P28885	hypothetical protein
11	403	4.6	3225	2 D81702	adherence factor T
12	398	4.6	1979	2 C71622	hypothetical protein
13	390	4.5	3724	2 T18427	hypothetical protein
14	386.5	4.5	3394	2 T18501	hypothetical protein
15	370	4.3	1711	2 T18429	hypothetical protein
16	359	4.1	1939	2 T18372	repeat organellar
17	356.5	4.1	4981	2 T18489	hypothetical protein
18	351	4.0	3835	2 A42771	hypothetical protein
19	348.5	4.0	2329	2 H81702	reticulocyte-binding
20	345.5	4.0	1127	2 T28317	adherence factor T
21	342	3.9	2819	2 A90551	ORF MSV156 hypothet
22	340	3.9	5005	2 P28884	conserved hypothet
23	337	3.9	2485	1 H71621	hypothetical protein
24	330.5	3.8	3844	2 T18402	asparagine/asparta
25	330	3.8	3255	2 G81702	adherence factor T
26	327.5	3.8	4550	2 T18440	hypothetical protein
27	327	3.8	2910	2 T28156	DNA-directed RNA p
28	318.5	3.7	1701	2 A26868	major merozoite su
29	318	3.7	3973	2 B71612	hypothetical protein

RESULT 1

I40884

cytoxin L - Clostridium sordellii

C/Species: Clostridium sordellii

C/Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999

C/Accession: I40884

R/Green, G.A.; Schue, V.; Monteil, H.

Gene 161, 57-61, 1995

A/Title: Cloning and characterization of the cytoxin L-encoding gene of Clostridium

A/Reference number: I40884; MUID:95369733; PMID:7642137

A/Accession: I40884

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-2364 <RES>

A/Cross-references: EMBL:X82638; NID:gl0000694; PIDN:CAA57959.1; PID:gl0000695

S/Superfamily: cpl repeat homology

C/Keywords: cytotoxin

Query Match 99.6%; Score 8646; DB 2; Length 2364;
Best Local Similarity 99.8%; Pred. No. 1.2e-303;
Matches 1696; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	1	MNLVNAQLQKQWVYKFRIQEDEVYVAILNALAEYHNHNSSESVVEKYLKLDINNLTDNYL	60
Db	1	MNLVNAQLQKQWVYKFRIQEDEVYVAILNALAEYHNHNSSESVVEKYLKLDINNLTDNYL	60
Qy	61	NTYKSGRNKALKKFKYEYLTMEVLEKNNSLTPVEKNLHPIWTGGQINDTAINYNQWKD	120
Db	61	NTYKSGRNKALKKFKYEYLTMEVLEKNNSLTPVEKNLHPIWTGGQINDTAINYNQWKD	120
Qy	121	VNSDVTVKFYDSNAFLINTLTKKTIYESATNNTLESFRENLDNPFEDYKFKRMEIY	180
Db	121	VNSDVTVKFYDSNAFLINTLTKKTIYESATNNTLESFRENLDNPFEDYKFKRMEIY	180
Qy	181	DKQKHFIDYKSIQIENPEFIIDNIKTLYLSNEYSKDLALNKYIEESLNKITANNNDI	240
Db	181	DKQKHFIDYKSIQIENPEFIIDNIKTLYLSNEYSKDLALNKYIEESLNKITANNNDI	240
Qy	241	RNLKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKS	300
Db	241	RNLKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKS	300
Qy	301	INKPDSITNTSWEMIKLEAIKMYKEIPGYTSKGFMDLDEEVQSFESALSSKDSKSEIF	360
Db	301	INKPDSITNTSWEMIKLEAIKMYKEIPGYTSKGFMDLDEEVQSFESALSSKDSKSEIF	360
Qy	361	LPDDIKVSPLEVKIAPANNVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE	420
Db	361	LPDDIKVSPLEVKIAPANNVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE	420
Qy	421	GTDFTNTMKIFSDKLASINEDNMFMKITYLNLYKVGFPADVRSTINLSGPGYTGAYQD	480
Db	421	GTDFTNTMKIFSDKLASINEDNMFMKITYLNLYKVGFPADVRSTINLSGPGYTGAYQD	480

ALIGNMENTS

30	317.5	3.7	1726	1	SAZQGM	major merozoite su
31	317.5	3.7	1726	2	A45948	major merozoite su
32	317.5	3.7	2599	2	P90608	ABC transporter pe
33	316.5	3.6	1701	2	A54498	major merozoite su
34	316	3.6	1447	2	F82909	hypothetical prote
35	314.5	3.6	2496	2	A71616	secreted protein p
36	314	3.6	1819	2	D97033	uncharacterized pr
37	313	3.6	1306	2	T28313	ORF MSV152 probabl
38	311	3.6	1302	1	JC6009	surface-located me
39	310.5	3.6	1830	2	E82909	conserved hypothet
40	309.5	3.6	2244	2	F90563	hypothetical prote
41	309	3.6	1639	2	S05603	major merozoite su
42	306.5	3.5	1516	2	E71619	RAD2 endonuclease
43	306.5	3.5	2136	2	A05037	hypothetical prote
44	305.5	3.5	1957	2	T38077	hypothetical coile
45	303.5	3.5	1802	2	G71616	hypothetical prote

Db 421 GTDFNTWKIFSDKLAISNEDNMFMKITYNLKVGAFDVRSTINLSGPGVYTGAYQD 480
Qy 481 LLMFKDNSTNHLPELRLNPEFPKTKISQTEOETSLWSFNQARAKSOFEEYKKGYPE 540
Db 481 LLMFKDNSTNHLPELRLNPEFPKTKISQTEOETSLWSFNQARAKSOFEEYKKGYPE 540
Qy 541 GALGEDNDLPAQNTVLDKQVSKILSSMKTRNKEYIHIVQLQGDKIISYEASCNLFPSK 600
Db 541 GALGEDNDLPAQNTVLDKQVSKILSSMKTRNKEYIHIVQLQGDKIISYEASCNLFPSK 600
Qy 601 DPYSSILYQKNIEGSETAYYYVADAEIKEDKYRIPYQISNKNENIKLTFIGHCKSEFNT 660
Db 601 DPYSSILYQKNIEGSETAYYYVADAEIKEDKYRIPYQISNKNENIKLTFIGHCKSEFNT 660
Qy 661 DTFANLDOVDSLSSIEIETILNLAADISPKYIEINLLGCNMFYSIYAEETYPGKLLKIK 720
Db 661 DTFANLDOVDSLSSIEIETILNLAADISPKYIEINLLGCNMFYSIYAEETYPGKLLKIK 720
Qy 721 DRVSELMPSISQDSITVSANQYEVRIINEBGRKREILDHSGKWINKESIIKDISKEYISF 780
Db 721 DRVSELMPSISQDSITVSANQYEVRIINEBGRKREILDHSGKWINKESIIKDISKEYISF 780
Qy 781 NPKNKILVSKYLHELSTLLQEIERNNANSDDIDLEKKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPKNKILVSKYLHELSTLLQEIERNNANSDDIDLEKKVMLTECEINVASNIDRQIVEGR 840
Qy 841 IEKAKNLTSDSINYIKNEFKLIESISDLYDLKHQGLDPSHFISFEDISKTENGFRIRF 900
Db 841 IEKAKNLTSDSINYIKNEFKLIESISDLYDLKHQGLDPSHFISFEDISKTENGFRIRF 900
Qy 901 INKETGNSIETETKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHEVNTLN 960
Db 901 INKETGNSIETETKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHEVNTLN 960
Qy 961 SAPFIQSLIEYNTTKESLSNLSVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 SAPFIQSLIEYNTTKESLSNLSVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDETID 1020
Qy 1021 LLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPRLRQIEAKIGIMAVNLTAATAIVT 1080
Db 1021 LLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPRLRQIEAKIGIMAVNLTAATAIVT 1080
Qy 1081 SALGIASGFSILLVPLAGISAGISPLVNNELILODKATKVIDYFKHLSLAETEGAFLLD 1140
Db 1081 SALGIASGFSILLVPLAGISAGISPLVNNELILODKATKVIDYFKHLSLAETEGAFLLD 1140
Qy 1141 DKTIIMPODDLVSIEDFNNSNITLCKEIIWRAEGSGHTLTDDIDHPFSSPSITYRKPL 1200
Db 1141 DKTIIMPODDLVSIEDFNNSNITLCKEIIWRAEGSGHTLTDDIDHPFSSPSITYRKPL 1200
Qy 1201 SIYDVLNKKKIDFSDKDLMLVLPNAPNRFVGYENGWTPGFRSLDNDGTKLLDRIDHYEG 1260
Db 1201 SIYDVLNKKKIDFSDKDLMLVLPNAPNRFVGYENGWTPGFRSLDNDGTKLLDRIDHYEG 1260
Qy 1261 QFYWRYFAFTADALITKLKRYEDTNVRINLDGNTRSFIYVPVITTEIQIRKLSYFVGS 1320
Db 1261 QFYWRYFAFTADALITKLKRYEDTNVRINLDGNTRSFIYVPVITTEIQIRKLSYFVGS 1320
Qy 1321 GSYSLSPYNNMIDLAVENDTWIVDQVNVKNITTESDBEIOKGBELIENILSKLNIEDN 1380
Db 1321 GSYSLSPYNNMIDLAVENDTWIVDQVNVKNITTESDBEIOKGBELIENILSKLNIEDN 1380
Qy 1381 KIILNHTNIPYGDINESNRPISUTFSILEDINIIIEIDLVSYSKILLSGNCMKLIENS 1440
Db 1381 KIILNHTNIPYGDINESNRPISUTFSILEDINIIIEIDLVSYSKILLSGNCMKLIENS 1440
Qy 1441 SDIOCKIDHGFNGEHOXIFYSYIDNETKNGPIDYSKKEGFTAEFSNESIIRNIYMP 1500
Db 1441 SDIOCKIDHGFNGEHOXIFYSYIDNETKNGPIDYSKKEGFTAEFSNESIIRNIYMP 1500
Qy 1501 DSNLFTYSSKDLKDIRINKGDVKLLIGNYFKDDMKVLSFTIEDTNTIKLNGVYLDEN 1560
Db 1501 DSNLFTYSSKDLKDIRINKGDVKLLIGNYFKDDMKVLSFTIEDTNTIKLNGVYLDEN 1560

Qy 1561 GVAQILKFMNNAKSALNTSNSLMNFELESINIKNIYNNLDPNIEFILOTNFIISGNSIG 1620
Db 1561 GVAQILKFMNNAKSALNTSNSLMNFELESINIKNIYNNLDPNIEFILOTNFIISGNSIG 1620
Qy 1621 QFELICDKDKNIOPIYFINKFKIKETSYTYLVGNRQNLIVPEPSYHLDDSGNISSTVINFSOK 1680
Db 1621 QFELICDKDKNIOPIYFINKFKIKETSYTYLVGNRQNLIVPEPSYHLDDSGNISSTVINFSOK 1680
Qy 1681 YLYGDRYVNVKVIIAPNLYT 1700
Db 1681 YLYGDRYVNVKVIIAPNLYT 1700

RESULT 2

S10317

toxin B - Clostridium difficile

C:Species: Clostridium difficile

C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 15-Oct-1999

C:Accession: S10317; S21894; S22434

R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.

Nucleic Acids Res. 18, 4004, 1990

A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.

A:Reference number: S10317; MUID:90326540; PMID:2374729

A:Accession: S10317

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-2366 <BAR>

A:Cross-references: EMBL:X53138; NID:G40442; PIDN:CAA37298.1; PID:G40443

R:Michel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, X.

submitted to the EMBL Data Library, July 1991

A:Description: Comparative analysis of Clostridium difficile toxins A and B.

A:Reference number: S21894

A:Accession: S21894

A:Molecule type: DNA

A:Residues: 1271-2366 <EIC>

A:Cross-references: EMBL:X60984; NID:G40445; PIDN:CAA43299.1; PID:G40446

R:Von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, X.

Mol. Gen. Genet. 233, 260-268, 1992

A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.

A:Reference number: S22434; MUID:92293124; PMID:1603068

A:Accession: S22434

A:Molecule type: DNA

A:Residues: 1791-2366 <VON>

A:Cross-references: EMBL:X60984

C:Genetics:

A:Gene: toxB

C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin

Query Match 77.6%; Score 6733.5; DB 2; Length 2366;

Best Local Similarity 76.4%; Pred. No. 8.4e-235;

Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

Qy 1 MNLVNAQLOKMYVYKRIQDEYVAIILNALSSEYRNSESSVVEKYLKLDKINNTDNYL 60

Db 1 MSLVNRKQLEKMANVFRFTQDEYVAIILNALSSEYRNSESSVVEKYLKLDKINNTDNYL 60

Qy 61 NTYKSGRNKALKKPKFVLTWELKNNSLTPVEKNLHFTWIGGOINDTAINYNOMKD 120

Db 61 DTYKSGRNKALKKPKFVLTWELKNNSLTPVEKNLHFTWIGGOINDTAINYNOMKD 120

Qy 121 VNSDYTVKPYVDSNAFLINTLTKTIVESATNTNLTLESFRENLDNDFDYNNKFKRMEIY 180

Db 121 VNSDYTVKPYVDSNAFLINTLTKTIVESATNTNLTLESFRENLDNDFDYNNKFKRMEIY 180

Qy 181 DKQKHFDYKSOIEENFEIIDNITKYLSEYKDLKALNKYTEESLNKITTANGNDI 240

Db 181 DKQKHFDYKSOIEENFEIIDNITKYLSEYKDLKALNKYTEESLNKITTANGNDI 240

Qy 241 RNLEKFAEDLRLVNLQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFS 300

Db 241 RNLEKFAEDLRLVNLQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFS 300